基于卷积神经网络的实时环境光遮蔽计算 *

郭雨潇 [1]，陈雷霆 [1, 2, 3]

(1. 电子科技大学 计算机科学与工程学院，成都 611731; 2. 电子科技大学 广东电子信息研究院，广东 东莞 523808; 3. 广东省卫生厅 政务服务中心，广州 510060)

摘　要：环境光遮蔽是一种常用于计算机游戏以及可视化领域的低频全局光照模拟算法。已有的基于屏幕空间的环境光遮蔽计算方法虽然保证了计算的实时性，但是存在估计失真以及细节丢失等难以解决的问题。针对该问题，提出了一种结合低频光线追踪采样以及蒙特卡罗去噪的算法框架对环境光遮蔽进行实时计算。为了解决传统蒙特卡洛去噪算法无法实时处理的问题，提出了一种基于卷积神经网络的蒙特卡罗去噪算法，并针对问题对网络结构进行了改进和优化。验证实验证明基于卷积神经网络的方法能够对环境光遮蔽的去噪问题进行准确的处理，同时对卷积网络的改进在保持精度的基础上显著地提高了计算效率。对比实验显示了所提算法在保持与高频采样光线追踪算法相近效果的前提下可达到与基于屏幕空间环境光遮蔽计算方法相近的每秒数百帧计算效率。

关键词：环境光遮蔽；蒙特卡罗采样去噪；卷积神经网络；屏幕空间；自动编码器

中图分类号：TP391.41　　doi: 10.3969/j.issn.1001-3695.2018.01.0054

Convolution neural network based real-time ambient occlusion computing

Guo Yuxiao[1], Chen Leiting[1, 2, 3]

(1. *School of Computer Science & Engineering*, *University of Electronic Science & Technology of China*, *Chengdu* 611731, *China*; 2. *Institute of Electronic & Information Engineering of Dongguan*, *University of Electronic Science & Technology of China*, *Dongguan Guangdong* 523808, *China*; 3. *Government Affairs Service Center of Guangdong Provincial Health Department*, *Guangzhou* 510060, *China*)

Abstract: Ambient occlusion is widely used in real time rendering for computing visibility of ambient light, especially in computer game and visualization domains. Although existing screen-space solutions can provide real time performance, they cannot generate accurate results and preserve details. This paper proposed a new solution based on Monte-Carlo ray tracing. To achieve real time performance, this method sampled the visibility of each pixel with sparse rays and then denoised the ray tracing results with a convolutional neural network (CNN) . To meet the real-time requirement of ambient occlusion computing, the algorithm further improved and optimized the network structure. Experimental results show that proposed method not only achieves the real time performance but also produces more accurate and detailed results.

Key words: ambient occlusion; Monte Carlo denoise; convolution neural network; screen space; auto encoder

0　引言

环境光遮蔽是图形学领域中一种经典且常用的低频全局光照模型，由 Landis[1]在 2002 年提出。其计算方法为对于物体表面上任意一点，统计以该点为中心一定半径内的上半球面区域中的遮挡区域与非遮挡区域的比例，从而获得该点受到全局光照强度的强弱。由于该模型计算方法简单且效果逼真，常被应用于计算机游戏以及计算机可视化等领域。传统地，获得场景中的几何遮挡关系需要结合光线追踪以及蒙特卡罗采样法对三维空间进行数百次采样，计算过程需要耗费大量的资源和时间。为了将环境光遮蔽应用于实时渲染的场景中，Mittring [2]提出了基于屏幕空间的环境光遮蔽计算方法。该方法首先通过光栅化渲染管线获得屏幕空间的像素深度信息，与一定半径范围内的相邻像素计算遮挡关系，从而获得近似的环境光遮蔽计算结果。通过在二维的屏幕空间进行采样，极大地节约了计算开销，使得实时计算环境光遮蔽成为可能。针对其效果有限的问题，随后 Bavoil 等人[3]的研究引入了屏幕空间的法向量，使得采样过程更加准确。Bavoil[4]和李文耀等人[5]为了解决失真和细节丢失等问题，引入了多尺度环境光遮蔽计算方法。本质上讲，基于屏幕空间是利用一个光栅化的二维投影平面结果估计对应

收稿日期：2018-01-08；**修回日期：**2018-03-09　**基金项目：**国家“863”计划资助项目（2015AA016010）；广东省应用型科技研发专项资金项目（2015B010131002）；东莞市重大科技项目（2015215102）

作者简介：郭雨潇（1989-），男，四川成都人，博士研究生，主要研究方向为实时渲染、深度学习（yuxiao.guo@outlook.com）；陈雷霆（1967-），男，教授，博导，博士，主要研究方向为虚拟现实、计算机图形学．

的三维空间结构和关系，进而计算对应的环境光遮蔽结果。该问题是一个典型的病态问题。一些相似地基于场景理解和分析的工作表明[6,7]，解决该类问题仍然是具有挑战性的。另外，该问题对实时性要求进一步限制了所用方法的计算开销。因此，基于屏幕空间的算法难以避免地会存在估计不准确和失真等问题。

本文的研究基于一个重要的观测，基于光线追踪的计算方法即使在采样率有限的情况下，虽然包含了大量的采样噪声，亦能较为准确地表达和估计大体的环境光遮蔽结果。如果能够将其包含的噪声进行剔除，即可获得质量较高且对环境光遮蔽估计准确的计算结果。因此，本文将该问题转换为实时蒙特卡罗采样去噪问题进行处理。蒙特卡罗去噪问题是图像去噪中的一个研究分支，旨在对图形学渲染领域中使用蒙特卡罗采样方法引起的图像噪声进行去除和抑制。一些早期的关注该问题的研究包括 Rushmeier 等人[8]以及 Mccool[9]的工作，主要使用了与普通图像去噪问题相似的一些滤波和优化算法。随后，Sen 等人[10]发现解决该特定问题的关键在于结合渲染和采样过程中的一些特定信息对问题进行处理，如表面法向量以及深度图像等额外信息。在此基础上，Rousselle[11]和 Moon 等人[12]相继提出了通过多个联合滤波器对结果进行进一步优化的方法。针对多个联合滤波器的权重调整困难问题，Kalantari 等人[13]提出了使用一个多层感知机对系数进行自动调整的想法。该类研究与本文相比，一个重要的区别是算法对计算设备和响应时间的要求不同。传统的应用于视觉特效渲染的蒙特卡罗采样方法，受到光线间复杂相互作用关系的限制，只能使用 CPU 作为处理单元，渲染时间往往在数分钟以上。所以基于该类应用的蒙特卡罗去噪研究，采用 CPU 和 GPU 均可满足要求，对响应时长的要求在数秒内即可接受。对于实时环境光遮蔽计算，由于仅在物体表面做线面求交运算，均采用了 GPU 作为运算设备，一些典型的处理方法包括文献[14,15]。所以，针对该问题的去噪算法设计的重点在于算法的 GPU 兼容性以及实时响应能力。基于以上两点考量，现有的蒙特卡罗去噪算法难以满足本文应用场景的要求。

近年来，卷积神经网络被广泛地应用于图像分类[16]、目标检测[17]和语义分割[18]等计算机视觉任务中，并取得了优异的结果。其中，Ronneberger 等人[19]提出了 UNet 结构并引入了跳跃连接，用于解决图像回归问题。卷积神经网络与传统方法相比，具有自动从数据中获取特征和先验以及共享权重等特点，使其计算具有高效性。另外，基于 GPU 的卷积神经网络算法和框架[20]设计已经较为成熟，能够在 GPU 中进行端到端的计算。如果能够设计一种高效的卷积神经网络并将其引入到蒙特卡洛去噪中，即能够使得问题得到解决。

针对以上存在问题和观察，本文提出了一种实时的蒙特卡罗采样去噪算法，对环境光遮蔽的实时计算问题进行处理。方法以低采样率下基于光线追踪方法的环境光遮蔽计算结果为输入，通过卷积神经网络对无噪声的结果进行回归和预测，以高质量的环境光遮蔽结果为输出。同时，针对实时性的要求，对卷积神经网络结构进行了进一步地优化，提出双路独立编码和选择跳跃连接等改进。另外，对卷积神经网络的参数和优化方法也进行详细的讨论。为了训练目标网络，从场景数据集中通过物理渲染的方法获得大量的训练和测试数据。

验证实验表明了基于卷积神经网络的方法求解该问题的有效性，同时说明了针对该问题对网络结果和参数设计所做改进的合理性。与基于光线追踪和屏幕空间的环境光遮蔽的对比实验表明，所提方法不仅具有光线追踪方法在高采样率下的准确度以及细节还原度，同时还达到了每秒数百帧的效率，满足了实时计算的要求。

1 算法流程

本文提出了一种全新的将基于光线追踪计算的低采样率下环境光遮蔽与蒙特卡罗去噪相结合的计算方法。其整体的算法流程如图 1 所示。算法的前半部分与传统使用光线追踪计算环境光遮蔽的方法相似。以场景的几何信息和待计算的视点为输入，首先通过视点像素与物体表面进行求交判断，获得像素点与物体表面的精确交点；随后对以各像素对应的物体表面点为中心的上半球面进行蒙特卡罗采样。与传统方法重复数百次采样过程不同，本文仅对各像素进行 4 次采样。随后将该初步计算的低采样率的环境光遮蔽计算结果作为蒙特卡罗去噪算法的输入，最后输出质量较高无噪声的环境光遮蔽计算结果。因此，算法的整体计算时间 T 由低频采样计算开销 T_R 和蒙特卡罗去噪开销 T_D 两部分组成。

$$T = T_R + T_D \tag{1}$$

基于光线追踪的蒙特卡罗采样部分是计算机图形学渲染领域中一个被广泛和深度研究的方向，其主要包含基于光学规则的光线生成、用于线面求交的碰撞检测算法以及加速空间查找的空间划分算法构成。由于本文的重心在于提出一种能实时对含噪环境光遮蔽结果进行去噪的算法，所以不在此对光线追踪的计算部分做过多讨论，使用了 Aila 等人[15]的方法和实现作为本文光线追踪采样部分的实现方法。接下来将对蒙特卡洛去噪部分进行详细的介绍和讨论。

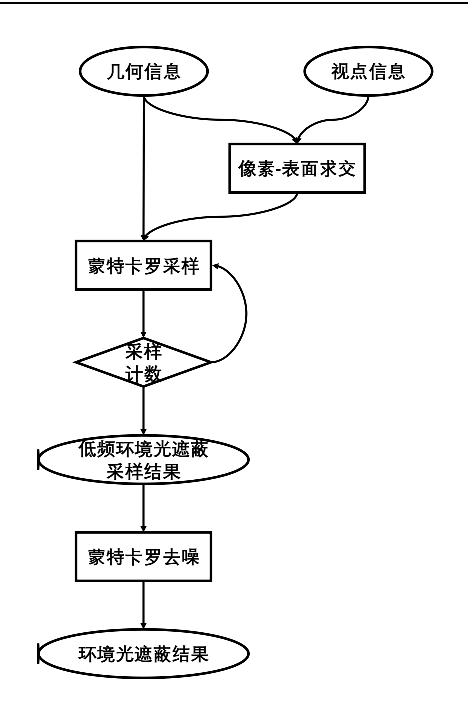


图 1　算法流程

2　蒙特卡罗去噪算法

2.1　问题定义

蒙特卡罗去噪拟解决的问题具有如下定义：

$$\hat{\theta}_p = \arg\min_{\theta} l\left(\bar{c}_p, g\left(X_p : \theta\right)\right) \tag{2}$$

其中：$X_p = \{c_p, f_p\}$，c_p 为低频采样下的渲染和计算结果，f_p 为对应的用于辅助求解的额外信息，通常使用深度或者表面法向量等，X_p 为算法的输入；$g(X:\theta)$ 为一个映射函数，θ 为该函数的参数；$\bar{c}_p$ 为对应的真实值。该问题通过提出相应的映射函数和函数对应的参数从而最小化误差，对问题进行处理。显然地，$\bar{c}_p$ 在实际使用时无法获得，一种常见的方法是使用一种或多种的滤波操作替代映射函数；另外一种方法是在 N 个训练样本中进行学习，最小化映射函数在 N 个样本中的误差。如果该输入样本能够覆盖实际使用时的样本空间，该优化函数可转换为

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N}\sum_{i=1}^{N} l\left(\bar{c}_i, g\left(X_i;\theta\right)\right) \tag{3}$$

在本文中使用了卷积神经网络对问题进行拟合。同时，生成了由大量仿真场景构成的训练样本对网络进行训练。另外，使用了深度图像作为额外特征对去噪过程进行辅助。

2.2　卷积神经网络结构设计和优化

设计的卷积神经网络总体结构如图 2 所示。接下来对网络的设计和细节做详细的介绍。

1）自动编码器　自动编码器是一种常用于求解图像回归问题的网络设计结构。其典型的结构如图所示。结构通常由编码器、特征提取以及解码器三个部分组成。首先，输入图像将通过一个编码器，该结构通常由数个编码单元构成。在编码过程中，特征图的尺寸缩减，但是特征图的个数增加，从而实现从局部到全局特征的抽取。在通过编码器后，特征图将进入连续的特征提取阶段。该阶段通常保持特征图的尺寸和通道数量不变，通过连续的卷积操作在全局尺度对特征进行加强和提取。在解码过程中，特征图通过解码单元后，尺寸增大，通道数量减少，使得全局特征逐步转换为局部细节。

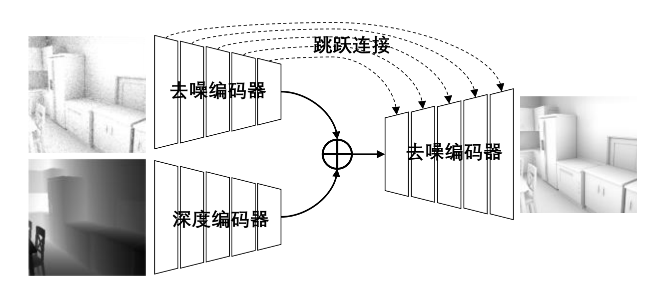


图 2　算法网络总体结构

2）编码单元结构　本文使用的编码单元结构如图 3 所示。其基本结构为两个连续地由卷积、批归一化（batch normalization）、线性整流函数（reLU）构成的小单元组成。第一个小单元的卷积采用了卷积核大小为 3，步长为 2，同时卷积输出通道为输入的 2 倍设置，负责将特征图尺度减小，特征通道增加。第二个小单元的卷积采用了卷积核大小为 3，步长为 1，卷积通道不变的设计，负责在特征尺度下对特征图进行特征提取和编码。

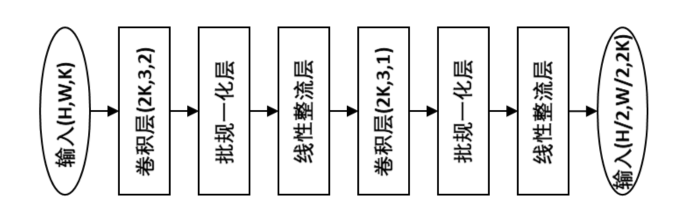


图 3　编码单元结构

3）解码单元结构　解码单元的输入包含由下层解码单元的输出以及同尺度编码器的输出两部分构成。首先，通过一个特征拼接操作将两部分的输入进行合并。引入了由反卷积、批规一化和线性整流函数等操作将特征图的尺度增加到原来的 2 倍，同时输出通道为下层解码单元输出通道的 1/2。最后，与编码阶段相似，使用了一个卷积核为 3，步长为 1 且卷积通道不变的卷积操作对特征进行解码操作。具体结构如图 4 所示。

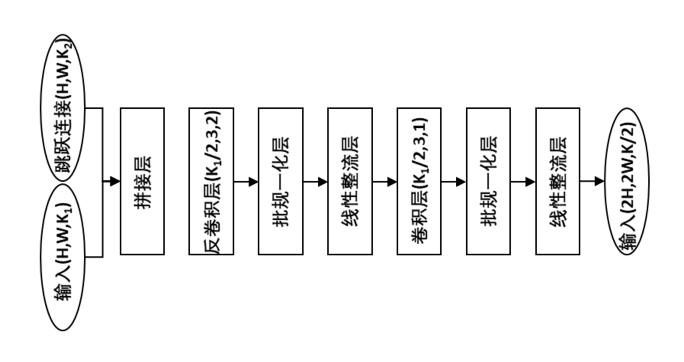


图 4　解码单元结构

为了进一步的优化网络计算效率，提升计算速度，本文还提出了双路独立编码和选择跳跃连接两种改进思想。

a) 双路独立编码。对于该问题，卷积神经网络需要同时接受含噪采样图像和深度图像作为输入。常用的处理办法是将两张输入图像进行拼接，通过一个共同的编码器结构。由于环境光遮蔽衡量了物体表面的局部可见性，所以过于细节的部分并不能直接决定环境光遮蔽的最终结果。基于以上考虑，提出了含噪图像和深度图像独立编码的结构。通过独立编码，在细节部分采用了独立的编码器，在包含了足够局部特征时对两路特征进行合并，随后进行特征提取和解码操作。双路独立编码使得单个分支的通道数量变为以前的 1/2，计算开销仅为共享编码结构的 1/4。通过减少细节阶段的含噪图像和深度图像的信息流动，提高了计算效率。

b) 选择跳跃连接。跳跃连接由 Ronneberger 等人[19]提出，是一种应用于卷积自动编码器的优化技巧。其核心在于将同特征尺度的编码器输出与解码器输入相连，使得编码阶段的信息能够直接传递到解码阶段，减小了信息的损失，提升了回归精度。一种常见和简单的处理方法是将所有编码器的特征进行拼接后直接作为解码器的输入。考虑到算法输入中含噪图像与输出结果的耦合度远高于深度图像与输出结果间的相关性，本文提出了一种选择跳跃连接方法，即仅将含噪图像分支的编码器与解码器进行跳跃连接。该方法减少了解码阶段输入通道的数量，使得反卷积操作的计算量减少了 1/3，提高了计算效率。

2.3　损失函数

使用卷积神经网络处理图像回归问题。常见的损失函数为 MSE（mean square error，平均平方误差）。Zhao 等人[21]发现，在卷积神经网络处理图像去噪问题中，即使使用 MSE 作为评价函数，MSE 损失函数的回归精度低于 MAE(mean average error，平均均值误差）和 SSIM（structural similarity，结构相似度)。其中 SSIM 损失函数具有如下定义：

$$L^{SSIM}(\bar{c},c)=1-\frac{2\mu_{\bar{c}}\mu_c+C_1}{\mu_{\bar{c}}^2+\mu_c^2+C_1}\cdot\frac{2\sigma_{\bar{c}c}+C_2}{\sigma_{\bar{c}}^2+\sigma_c^2+C2} \quad (4)$$

其中：μ 为均值；σ 为方差；特别地，$\sigma_{\bar{c}c}$ 代表真实值 $\bar{c}$ 和预测值 c 的协方差；C_1 和 C_2 为常数，用于维持数值稳定。通常，在使用时将图像分割为大小相等的数个小块，每个小块分别计算各自的误差，最后取算术平均作为总的预测误差。为了进一步探究不同损失函数在蒙特卡洛去噪问题中的性质，本文分别使用了三种不同的损失函数对网络进行训练，并且对最终结果进行了比较。

2.4　训练/测试数据生成

卷积神经网络的训练需要大量的标注数据作为训练样本。为了生成所需数据，本文选择了从仿真场景数据集中渲染环境光遮蔽结果和深度结果的方法进行数据生成。Song 等人[22]是一个包含约 50 000 左右仿真室内场景的三维模型数据集。对于每个场景，从其推荐的视角中选择得分最高的约 30 个视点进行渲染。对于环境光遮蔽结果，以 SPP（sample per pixel，逐点采样率）4 和 512 渲染，分别作为训练的含噪声的环境光遮蔽结果和无噪声的本征值。同时，相应的渲染相同视点对应的深度图像。本文选择了数据集中约 1 800 个场景，其中 1 500 个生成训练所用的数据，剩下 300 左右的场景则作为测试和评估结果所用的数据。这样，训练样本共计 46 000 左右，测试样本约为 9 000。

3　实验与结果分析

3.1　实验环境和实现细节

实验所用的硬件环境如下：一块 Intel i7 7700 CPU，一块 Nvidia GTX TITAN X 图形加速卡，主机内存 128 GB。

训练蒙特卡罗去噪算法所用的卷积网络训练平台为 MxNet，梯度更新算法采用了 AdaDelta 算法。在训练时每个迭代包含 100 张训练样本，共计训练了 100 000 次迭代。整个训练过程需要约 15 个小时，大约需要 3 GB 显存。对于，所有的算法，在本文生成的测试数据集中进行对比和验证。

3.2　卷积网络设计验证和分析

本文首先对采用的卷积网络结构进行了验证和分析。总体结果如表 1 所示。接下来将就各项进行具体分析。

表 1　卷积神经网络不同设计估计误差对比

设计细节	MSE/10^{-3}	时间/ms
本文方法+MAE	0.401	0.71
本文方法+MSE	0.455	0.71
本文方法+SSIM	0.413	0.71
本文方法+无深度图像	0.486	0.70
本文方法+卷积通道增加	0.397	1.37
本文方法+卷积通道减少	0.534	0.48
本文方法+共享编码器	0.398	0.93
本文方法+全跳跃连接	0.403	0.86

1）损失函数对结果的影响　在训练时，分别使用了 MSE、MAE 以及 SSIM 三种损失函数，其结果如表 1 中前三项所示。其中，MAE 作为损失函数的结果具有最好的数值结果，其次是 SSIM，最后是 MSE。值得注意的是，一般认为相同的损失函数和评价函数能够取得最一致的结果。本文采用了 MSE 作为评价函数，然而 MSE 作为损失函数的结果却并不如 MAE 和 SSIM，说明通过深度学习求解蒙特卡罗去噪问题与传统方法存在一定的区别。由于 MAE 具有最好的结果，本文其他网络均采用了 MAE 作为训练的目标损失函数。

2）深度图像对结果的影响　蒙特卡罗去噪问题中，使用额

外的信息对求解该问题具有重要的帮助。本文引入了深度图像作为额外信息，试图取得更好的结果。表 1 中对比了是否使用深度信息对最终结果的影响。为了确保卷积网络具有可比性，针对无深度图像的网络将其编码器的卷积通道数量变为原来的 2 倍。显然地，深度图像的引入能够获得更好的结果，证明了额外的信息对基于深度学习的蒙特卡洛去噪亦具有帮助。

3）卷积通道数量对结果的影响　卷积通道的数量决定了卷积网络的拟合能力和计算效率。较大的卷积通道一般具有更强的拟合能力和更大的计算开销，较小的卷积通道则反之。本文网络选择了卷积通道数目为 4。表中探讨了在本文拟解决问题中卷积通道数目分别为 2（表中卷积通道减少项）和 8（表中卷积通道增加项）时对结果的影响。相比与卷积通道为 2 的情况，本文设计具有显著地精度提升，但是牺牲了一定的计算效率。相比与卷积通道为 8 的情况，本文设计精度下降并不明显，但是计算更加高效。故本文通道的选取为 4 能够有效地平衡了精度和性能，是一个合理的选择。

3）双路独立编码对结果的影响　针对深度图像和含噪图像，使用了独立的编码器进行编码，有效地减少了编码阶段的连接数量，降低了计算开销。针对该设计，本文与共享编码器的网络进行了对比，该网络直接在输入时对深度图像和含噪图像进行拼接，同时将共享编码器的卷积通道调整为独立编码器的 2 倍。在数值误差上，两种网络结构具有近似的结果；但是在计算速度上，双路独立编码具有更优的结果，证明了双路独立编码结构在解决该问题上的有效性。

4）选择跳跃连接对结果的影响　在网络结构设计中，本文跳跃链接结构仅将含噪图像编码器与去噪解码器进行了连接。在此，本文亦讨论了同时将两路编码器和解码器进行连接的情况，结果如表 1 中全跳跃连接项所示。本文设计结构与全跳跃连接相比，具有几乎相同的计算精度，同时计算速度更快。本文认为由于含噪图像和网络输出结果间的耦合程度远远高于深度图像，所以在该情况下加入深度图像的跳跃连接对于结果提升不明显，反而由于增加了连接数量使得计算效率下降。

3.3 对比实验结果和分析

为了进一步地对本文方法进行验证，选取了一些具有代表性的环境光遮蔽算法进行对比实验。首先是基于光线追踪的方法，分别选取了 SPP（sample per pixel，每像素采样率）为 4、16 以及 64 的结果作为对比。随后，选取了 HBAO 和 $SSAO^{+}$两种具有代表性的基于屏幕空间的环境光遮蔽算法加入对照，其采样次数分别为 8、64 次。测试样本选用了本文生成的测试数据集，实验将从数值结果和视觉结果两部分进行比较和分析。

3.3.1 数值误差对比

首先，本文从数据上对比了以上方法，选择了 MSE 以及单帧处理时间两项指标对算法的性能进行衡量，如表 2 所示。从回归精度上比较，本文方法以 SPP 4 次的低采样光线追踪计算结果为输入，输出的数值误差与 SPP 64 的结果相近，具有较高的回归精度。与基于屏幕空间的方法相比较，显著地提高了计算精度和准确度。从处理时间上比较，本文方法包含低频结果渲染时间（2.78 ms）和卷积神经网络推理时间（0.71 ms），亦能达到每秒数百帧的处理速度，满足了实时运算的需要，而相同精度的 SPP 64 的光线追踪算法帧率仅为 20 帧左右。另一方面，虽然常见的基于屏幕空间的算法能够达到更高的处理速度，但是每秒数百帧的情况下，性能已经不在是算法追求的唯一标准，精度的提升反而更加重要。

表 2　不同方法数值误差和计算性能对比

方法	$MSE/10^{-3}$	计算时间/ms
光线追踪 SPP4	2.489	2.78
光线追踪 SPP16	0.734	11.09
光线追踪 SPP64	0.269	43.97
$SSAO^{+}$ SPP 16	1.274	0.97
HBAO SPP 64	0.979	1.34
本文方法	0.401	3.49（2.78+0.71）

3.3.2 视觉结果对比

同时，本文亦从视觉效果上对基准值/本文方法和基于屏幕空间的 HBAO 和 SSAO 算法进行了比较，如图 5 所示。

显然地，本文方法总体上具有更加准确的估计精度。具体地分析，本文方法在果盘（图 5 中蓝色方框所示，见电子版）、窗帘（图 5 中紫色方框所示）等几何结构较为复杂的区域能够较为准确地估计局部环境光遮蔽情况，而对比方法存在着估计不准确等情况。另外，在一些细小的物体估计中，如柜子的把手(图 5 中红色方框所示)、摇篮的栅栏(图五中黄色方框所示）等处，对比方法难以准确估计如此细小的物体，而本文方法则不存在此问题。视觉对比证明了本文方法在效果上具有比较明显的提高。

4　结束语

针对基于屏幕空间的环境光遮蔽估计方法存在的模糊和细节丢失问题，本文提出了一种将光线追踪和蒙特卡洛去噪结合的方法，利用卷积神经网络和大量训练数据对该问题进行改进。实验证明改进后的算法具有更好的精度，并且亦能够达到实时处理的要求和良好的扩展性。

然而，由于本文方法使用了基于光线追踪的蒙特卡洛采样环境光计算方法，而现有计算机游戏领域均使用了基于光栅化的渲染管线，使得无法直接应用在计算机游戏领域，这是本文下一步的研究重点。

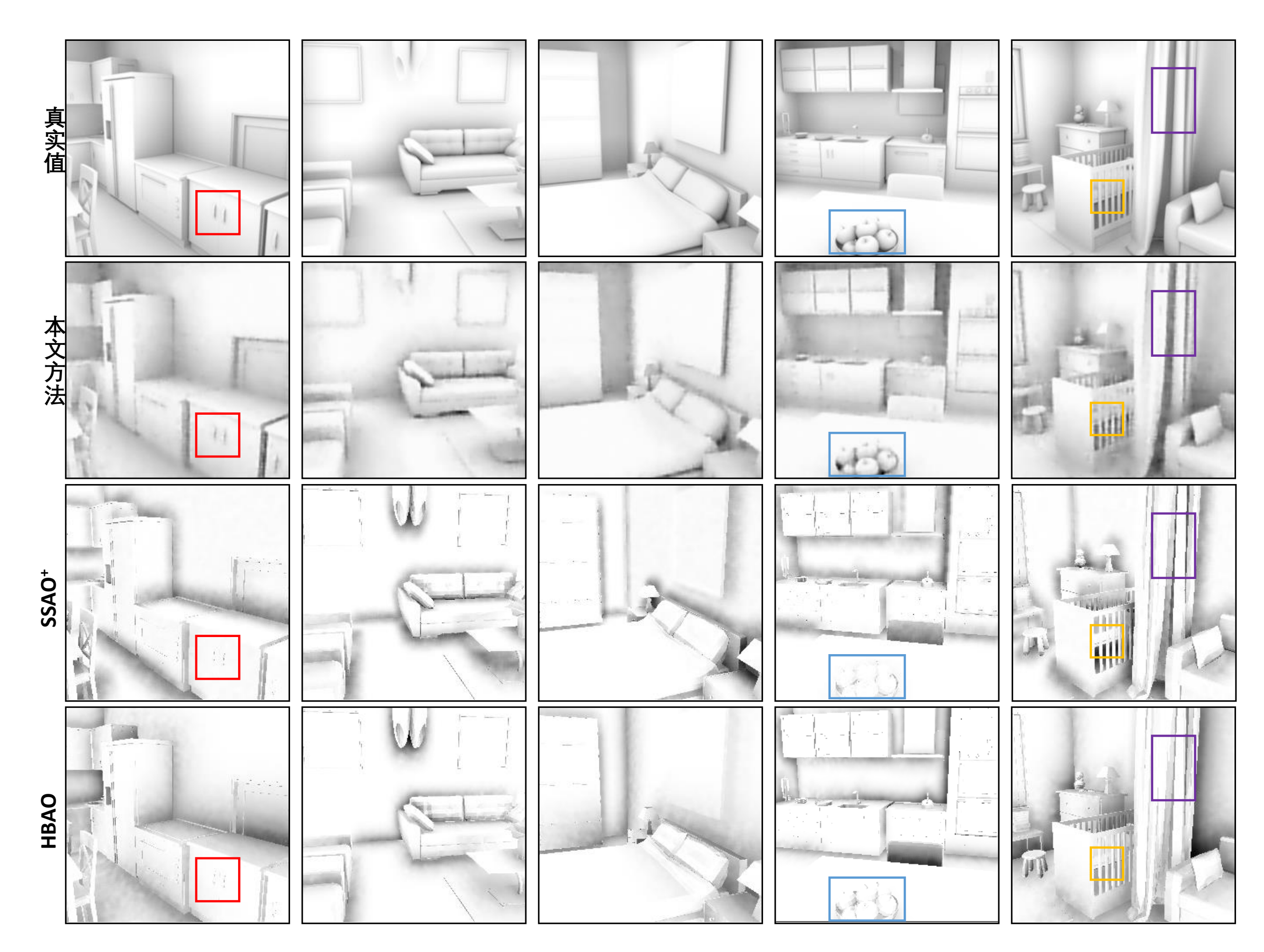


图 5 不同方法视觉对照结果

参考文献：

[1] Landis H. Production-ready global illumination [C]// Proc of the 35th ACM SIGGRAPH. New York: ACM Press, 2002.

[2] Mittring M. Finding next gen: cryengine 2 [C]// Proc of the 40th ACM SIGGRAPH. New York: ACM Press, 2007: 97-121.

[3] Bavoil L, Sainz M, Dimitrov R. Image-space horizon-based ambient occlusion [C]// Proc of the 41th ACM SIGGRAPH. New York: ACM Press, 2008.

[4] Bavoil L, Sainz M. Multi-layer dual-resolution screen-space ambient occlusion [C]// Proc of the 42th ACM SIGGRAPH. New York: ACM Press, 2009.

[5] 李文耀, 任重. 自适应的多层屏幕空间环境光遮蔽 [J]. 计算机辅助设计与图形学学报, 2011, 23 (8): 1294-1303.

[6] Yu Tianshu, Wang Ruisheng. Scene parsing using graph matching on street-view data [J]. Computer Vision and Image Understanding, 2016, 145: 70-80.

[7] Zhao Yibiao, Zhu Songchun. Scene parsing by integrating function, geometry and appearance models [C]// Proc of the 31st IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2013: 3119-3126.

[8] Rushmeier H E, Ward G J. Energy preserving non-linear filters [C]// Proc of the 21st Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 1994: 131-138.

[9] Mccool M D. Anisotropic diffusion for Monte Carlo noise reduction [J]. ACM Trans on Graphics, 1999, 18 (2): 171-194.

[10] Sen P, Darabi S. On filtering the noise from the random parameters in Monte Carlo rendering [J]. ACM Trans on Graphics, 2012, 31 (3): 1-15.

[11] Rousselle F, Manzi M, Zwicker M. Robust denoising using feature and color information [J]. Computer Graphics Forum. 2013, 32 (7): 121-130.

[12] Moon B, Jun J Y, Lee J H, *et al.* Robust image denoising using a virtual flash image for Monte Carlo ray tracing [J]. Computer Graphics Forum. 2013, 32 (1): 139-151.

[13] Kalantari N K, Bako S, Sen P. A machine learning approach for filtering Monte Carlo noise [J]. ACM Trans on Graphics, 2015, 34 (4): 122-134.

[14] Tang Min, Zhao Jieyi, Tong Ruofeng, *et al.* GPU accelerated convex hull computation [J]. Computers & Graphics, 2012, 36 (5): 498-506.

[15] Aila T, Laine S. Understanding the efficiency of ray traversal on GPUs: [C]// Proc of the 1st Conference on High Performance Graphics. New York: ACM Press, 2009: 145-149.

[16] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Deep residual learning for image recognition [C]// Proc of the 34th IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2016: 770-778.

[17] Ren Shaoqing, He Kaiming, Girshick R, *et al.* Faster R-CNN: towards real-

time object detection with region proposal networks [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2017, 39 (6): 1137.

[18] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2014, 39 (4): 640.

[19] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [C]// Proc of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2015: 234-241.

[20] Sze V, Chen Y H, Yang T J, *et al.* Efficient processing of deep neural networks: A tutorial and survey [J]. Proceedings of the IEEE, 2017, 105 (12): 2295-2329.

[21] Zhao Hang, Gallo O, Frosio I, *et al.* Loss functions for image restoration with neural networks [J]. IEEE Trans on Computational Imaging, 2017, 3 (1): 47-57.

[22] Song Shuran, Yu F, Zeng A, *et al.* Semantic scene completion from a single depth image [C]// Proc of the 34th International Conference on Computer Vision and Pattern Recognition. Pscataway, NJ: IEEE Press, 2016: 1746-1754.